December 1999

THERdbASE v1.2

Tutorial Introduction

Appendix B

Human Exposure Modeling in THERdbASE
A Tutorial Introduction

Edwin J. Furtaw, Jr., William H. Engelmann, Larry C. Butler, and Joseph V. Behar

1. HUMAN EXPOSURE MODELING

Human exposure modeling is the science of assessing, by means of computational models, the pollutant exposures of human beings. This multi-disciplinary science involves considering what environmental media people contact, where and for how long they contact those media, and how much pollutant from those media is transferred to the human organism. In the fullest sense, human exposure modeling concerns itself with the full range of pollutant movement from source of input into the environment, all the way to various organs or tissues within the human organism, where health effects may occur. Thus human exposure modeling considers a continuum from "source-to-dose." The science of Risk Assessment then couples the results of exposure assessment or exposure modeling with toxicological data to estimate probabilities of certain health effects which may be caused by those pollutants. *Therefore, human exposure modeling is a necessary component of Risk Assessment.*

The terms "exposure modeling" and "exposure assessment" are nearly synonymous. In "exposure modeling" the user explicitly uses computational models to conduct exposure assessments. "Exposure assessment" is somewhat broader and may include exposure-related data analysis that is not considered to be formal modeling.

1.1. General overview

This section provides a global overview of modeling, human exposure modeling, and the use of THERdbASE (Total Human Exposure Risk database and Advanced Simulation Environment) for conducting human exposure modeling.

1.1.1. Global Philosophy of modeling

In a general context, "modeling" refers to the creation or use of conceptual or symbolic systems to represent something about the real world. Models can be purely conceptual, or they can be embodied in physical objects, or in the form of mathematical equations, algorithms, or computer programs which perform calculations.

Since models only represent real-world phenomena, as distinguished from actually being the objects or phenomena they represent, all models are, therefore, imperfect and incomplete. Reality seems so complex and intricate that most models of real systems are vast oversimplifications of the actual systems. These simplifications are frequently made as assumptions inherent within the design of the model. These simplifying assumptions are made for the expediency of allowing us to prescribe and perform calculations which we are capable of understanding given the limitations of human knowledge itself. To use an analogy, a model is like a picture of an object, which is quite different than the object itself.

Keep in mind this inherent shortcoming of all models. Modelers should remember that the results of modeling are subject to error and inaccuracy. Models simply may not give an accurate representation of the real world. Hence a healthy dose of skepticism should be applied when interpreting model outputs. The wise modeler will always look for ways to use reliable empirically measured data to evaluate the validity of model predictions. In fact, the design of models should reflect the ability of actual measurements to characterize a system. If something cannot be reliably observed and measured, it probably is not wise to rely on a conceptual model of that thing, because we have no way of testing the validity of the model. An untested model is the equivalent of an untested hypothesis. As interested model offers more of a question than a meaningful answer.

However, despite the limitations of all thinking and modeling, models can assist us with practical tasks such as assessing the likely consequences of certain actions. Human exposure modeling is one such practical use of modeling. As described below, human exposure modeling allows us to estimate the effects which may be caused on human health by pollutants in the environment.

1.1.2. Overview of human exposure modeling

Human exposure modeling deals with the behaviors of both the environment and human beings within the environment. As we humans go about our daily lives, the activities we engage in and the locations we occupy inevitably bring us into contact with pollutants that reside in and transfer among the various environmental media. Hence the movements of both pollutants and humans within the environment are subjects of interest to the exposure modeler. Furthermore, after pollutants contact humans, the pollutants enter and move around within the body as they become pharmacokinetically distributed, metabolized, and eliminated from the body. This disposition within the body is also considered to be within the realm of human exposure modeling. Thus, human exposure modeling is a highly interdisciplinary science, encompassing the fields of environmental science, sociology and human behavior, physiology and biology, analytical chemistry, as well as the mathematical, statistical, and computer sciences needed to create, use, and analyze outputs from the models.

Because of the complexity and wide range of considerations within the scope of human exposure modeling, different aspects of the science have evolved at different paces and to different levels of sophistication. For example, the science of atmospheric pollution modeling has evolved through several generations of software development pursued by many hundreds of researchers. By contrast, there have been only a few large-scale studies of human activity patterns specifically oriented toward understanding the impact of personal behavior on pollution exposure. Given the diversity of developments of the various constituent sciences that contribute to human exposure modeling, the information and resources needed by exposure assessors is scattered and available, if at all, from widely diverse sources and in different degrees of development. Hence the resources needed for exposure modeling are obtained in a somewhat piece-meal way by the assessor. This makes the utilization of exposure models an inherently modular process. The modules or pieces of information needed by an exposure assessor include numerous and diverse databases and sub-models that deal respectively with the numerous pollutant sources, environmental media, pathways by which humans contact these media, human behavioral elements, and physiological and pharmacokinetic aspects of humans.

After the modeler defines the types of data and models needed for a given assessment, the next and possibly most daunting task is to identify and obtain available information resources to fill those needs. If the needed resources are available, they can be acquired and, usually after screening, modification, and manipulation, can be incorporated into an overall assessment. Thus, pieces of an exposure assessment, in the form of databases and models, are assembled by the assessor or modeler. Using these assembled data and models, the modeler then performs the functions needed to arrive at the desired assessment. The assembly of the needed pieces is generally a somewhat informal procedure, because the various data and modeling pieces are not usually software-compatible for deployment in a single software system. This significantly increases the amount of time and effort that must be expended in conducting an exposure assessment. Various data are extracted from their respective databases and prepared as needed for each part of the model. The various sub-models are executed in their respective softwares. Frequently, the output from one sub-model will be required as input to another sub-model. The burden generally falls on the modeler to be able to convert one output to another input. Thus exposure modeling can become a laborious project requiring the assessor to obtain and become proficient in a large number of databases, analysis tools, and modeling platforms. Clearly, a software system which assists the modeler in assembling and organizing the pieces needed into an overall assessment would have considerable value to the modeler. *THERdbASE is one such software system because THERdbASE combines diverse databases and models in one convenient software platform..*

1.1.3. Human exposure modeling in THERdbASE

The Total Human Exposure Risk database and Advanced Simulation Environment (THERdbASE) is a software system which combines in one user-friendly platform many of the types of databases and models needed for human exposure assessment. THERdbASE is a Windows application program for personal computers which was developed in a Cooperative Agreement between the Environmental Protection Agency (EPA) and the University of Nevada-Las Vegas.

When a user opens the THERdbASE software, a virtual desktop is shown. The basic organization of this desktop is that databases are listed on the left half and models are listed on the right side. A user can select from a large number of databases, which are arranged in a data tree structure. The branches of the data tree represent different categories of databases frequently used in exposure assessment. The databases are organized into ten (10) categories as follows:

1. Demographics
2. Human Activity Patterns
3. Food Consumption Patterns
4. Food Contamination
5. Chemical Properties
6. Pollutant Source Information
7. Environmental Characteristics
8. Environmental Contamination Levels
9. Physiological Parameters
10. User-created databases.

The number of databases currently available in THERdbASE is too large to show all the databases on the desktop at one time. Hence, the user moves databases onto and off from the desktop as needed. There are numerous database functions which users can perform on the databases in THERdbASE. These functions are described in detail in the appropriate sections of the User Manual.

The number of models currently available in THERdbASE is eleven (11). This is far fewer than the number of databases that will operate in THERdbASE. All the models in THERdbASE are listed on the right half of the virtual desktop, so the user does not need to move models to and from the desktop as must be done for databases. The user can quickly scan the list of models to see what is available.

One of the great advantages of using THERdbASE for exposure modeling is the ease with which models are deployed. In THERdbASE, all the models have similar user interfaces. Also, the models are programmed to automatically obtain their needed data from the databases within THERdbASE. Model outputs are automatically written as data files which appear in the same format as all the other THERdbASE databases. Hence the process of exposure modeling can be greatly facilitated and streamlined with THERdbASE.

THERdbASE is an incomplete work-in-progress. Not all the types of data and models needed for thorough exposure modeling are currently available in THERdbASE. As time goes on, the THERdbASE developers plan to make available through THERdbASE a much wider range of data and models, in order to eventually achieve the goal of providing users with one easy-to-use package containing all the data and models needed for truly total human exposure assessments.

The models available in THERdbASE comprise a very useful subset of models for conducting human exposure assessments. These include models in several categories including indoor air pollutant emission source models, indoor air pollutant concentration models, a human activity pattern simulation model, inhalation exposure and dose models, and a combined inhalation and dermal exposure and dose model. These models fall into the five general categories of exposure-related models, as discussed below in Section 1.2.

To run one of the exposure-related models in THERdbASE, the user follows a simple 4-step procedure, as described below.

1. ***Select desired model***. The user selects a model for deployment by simply highlighting the model by clicking on the model name in the list on the right-hand side of the THERdbASE desktop. When a model name is highlighted, the user can view and read information or metadata about that specific model by clicking on the "I" (information) icon. The metadata file for each model gives a description of that model's origin, inputs, outputs, how it works, applicability, references, and other pertinent information.

2. ***Set model inputs***. When the Input icon (gears with arrow pointing inward) is clicked, a dialog box is opened. Here the user enters a name for the model input settings which the user wishes to run. The user then must specify the data to be used for each input field of that model. THERdbASE facilitates this data specification process by showing a list of the required inputs. The user selects each item on the list and THERdbASE provides a dialog box within which the user enters or selects the desired values. For many parameters, the user can select from among default or user-specified constant values, or from mathematically defined distributions of values, or from specified data files which contain distributions of the desired parameter values. These various types of data inputs are discussed in more detail in Section 1.4 below.

3. ***Run the model***. After creating a named set of inputs for a model, the user can run the model with those settings by simply clicking on the Run (gears) icon, then selecting the desired setting name. The model will run while showing a moving bar graph of an estimate of percent completion.

4. ***Review model outputs***. When the modeling run is completed, the user can review the model outputs or results by clicking on the Output icon (gears with arrow pointing outward), then selecting the named settings for the run to be reviewed. Outputs typically consist of several tabular and several graphical displays of calculated variables.

1.2. Model categories

Often within the field of human exposure assessment modelers conceive and describe the exposure process from the point of view of the pollutant as it moves from a source to human receptors where it may cause health effects. This exposure process is shown diagrammatically in *Figure 1* below. Models are categorized logically as to where they occur along the source-to-effects continuum. This way of categorizing models assists conceptually in understanding the complex relationships and cause-and-effect aspects among the various steps in the process.

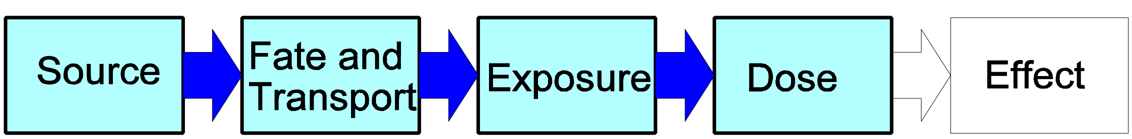

Figure 1. **The Exposure Process.**

The overall concept or paradigm for the exposure process is as follows: pollutants are released from some container into the environment. Within the environment, the pollutants move and distribute themselves into various environmental media (such as air, water, soil, and plants), where they may also be transformed in various ways. Humans then contact these media and some of the pollutants move into the human body. Within the body, pollutants (and possibly their metabolites) further distribute into various organs and tissues. These doses of pollutants to the various organs and tissues may cause health effects.

Models have been created to simulate this pollutant movement at various steps of the process. The models from the various source-to-dose categories are all used in a thorough exposure assessment. Risk assessment then carries the analysis to the next step by linking the results of the exposure assessment with an analysis of the risk (probability) of health effects which the exposure may cause. Hence, exposure assessment is one of the major sciences upon which risk assessment depends. Characteristics of models in each of the five major steps of the exposure and risk assessment process are discussed in the following sub-sections.

1.2.1. Source models

Source models characterize pollutant emissions or releases into one or more environmental media. Pollutants are conceived to originate at a "source." This term must be interpreted situationally. One view of a source may not be applicable for all situations. For example, if we consider sulfur dioxide (SO_2) emissions into the air from a power plant smokestack, that smokestack may be considered the source since that is the point of entry of pollutant into the environment (the atmosphere). But from another perspective, if one were to model the formation of SO_2 within the combustion process in the furnace that feeds that smokestack, then the modeler must look to the coal which contains sulfur as the source, rather than the smokestack. But if one were modeling the formation over geological time of the coal itself and its constituents, then that modeler would have to look even further back in time to understand the source of sulfur in the coal. Thus the concept of a pollutant source is relative. In the human exposure process, the term "source" generally implies the point of emission of pollutant from a man-made enclosure into some medium which is part of the open environment from which it can exchange with other compartments or media in the environment.

1.2.1.1. Overview

Source models in general should have several characteristics in order to be compatible with environmental media models. Essential information to be provided by source models are as follows:

1. They must identify what pollutants are emitted, and possibly in what form (such as vapor or particulate).
2. They must specify the medium or media into which they are emitted (such as indoor air, outdoor air, or a body of water). Sometimes it is important that the exact geographic location of the emissions be specified.
3. They must identify the rate of emission as a function of time, with start and stop times. The rate is specified as mass of pollutant per unit time.

1.2.1.2. List of source models in THERdbASE

The following source models are currently available in THERdbASE. The user can obtain information about each of these models by accessing the metadata for the model of interest.

1. Model 103: Source - Timed Application
2. Model 104: Source - Instantaneous Application

1.2.2. Environmental fate and transport (environmental media concentration) models

This category of model estimates the concentrations of pollutants in various environmental media.

1.2.2.1. Overview

Environmental fate and transport models are usually the most complex of the various model categories used in exposure and risk assessment. These models simulate the movement and possibly also the transformation of pollutants in multiple environmental media. Some of the most important environmental media that are considered in these models are:

1. outdoor air
2. particles suspended in outdoor air
3. indoor air in a single zone or in multiple rooms or zones
4. particles suspended in indoor air
5. rain droplets
6. water in a stream or other surface body of water
7. groundwater in an aquifer
8. sediment in a stream bottom
9. soil
10. dust particles which have settled onto a surface
11. food crops growing in a field
12. treated water at the tap in a consumer's home
13. prepared foods and beverages as consumed by people
14. sofa cushions which may adsorb pollutant vapor from indoor air (and may re-emit the pollutant back into the air later)
15. clothes which have been dry-cleaned
16. children's toys which may have pollutants adsorbed onto their surface

From this list, one can see that a very broad range of media are considered in environmental fate and transport models. These media range from large scale regional media such as the atmosphere and the oceans, to the very small-scale microenvironmental media such as the air in a shower stall when a person is showering or drinking a can of soda. In order to understand the movement of pollutants from their sources to human receptors, all these media must be considered in a thorough assessment.

Environmental fate and transport models should have several characteristics in order to be compatible with source models and human exposure models. Essential types of information to be provided by environmental fate and transport models are:

1. They must identify what medium or media are being considered. Sometimes it is important that the exact geographic location of the media be specified.
2. They must be able to accept inputs from source models. Sometimes it is important that the exact geographic location of the emissions be specified.
3. They must be able to handle pollutant interactions between media, such as the adsorption and desorption of an organic vapor between air and a solid or porous surface.
4. They must be able to output pollutant concentration in each specified medium as a function of time. The dimensions of the concentrations must be specified, and may have various forms, such as a mass per unit volume for water pollutants, a mole fraction expressed as parts per billion for an air pollutant, or a weight fraction expressed as parts per million or milligrams of pollutant per kilogram of soil.

1.2.2.2. List of environmental fate and transport models in THERdbASE

The following environmental fate and transport models are available in THERdbASE. The user can obtain information about each of these models by accessing the metadata for the model of interest.

1. Model 105: Indoor air - 2 Zones
2. Model 106: Indoor air - N Zones

1.2.3. Human exposure models

By definition, human exposure is the contact of a human with a pollutant. Human exposure models compute the extent of this contact.

1.2.3.1. Overview

Humans contact a wide variety of environmental media through the course of a normal day and throughout life. The most obvious such media include the air, the food and water and other beverages we ingest, and the surfaces and objects we touch. When these media are contaminated with pollutants, our contact with the media causes contact with the pollutants; hence exposure occurs.

Exposure has been described as the simultaneous occurrence at the same location of both a human and a polluted medium. Hence, exposure models must consider both the locations of people and their activities, because these factors determine the type and extent of human contact with various media. Environmental fate and transport models, described in preceding sections, provide concentrations of pollutants in the various media. Exposure is modeled by combining human location and activity patterns with pollutant concentrations which have either been measured or modeled with environmental fate and transport models.

Human exposure occurs via different pathways and routes. An *exposure pathway* refers to the conceptual sequence of locations and changes through which the pollutant moves as it makes its way from its source to a human receptor. Exposure route refers to the portal of entry of pollutant into the human body. Each exposure incident can be described by a pathway and route. For example, if I drink a glass of water containing a pollutant, the pathway describes how that pollutant got into the water, considering both its source and its environmental fate and transport through various stages of the hydrogeological cycle and water treatment and distribution processes. The exposure route in this example is ingestion, because the pollutant enters the human body by being taken in orally and swallowed into the digestive tract.

The three major routes of human exposure to environmental pollutants are ingestion, inhalation, and dermal absorption. There are other potential but less significant routes, such as injection and ocular absorption, but these are rarely considered explicitly in human exposure assessments, except possibly in pharmaceutical applications.

Different dimensions are used to characterize human exposure via different routes. Inhalation exposure has been defined in some references as the integral of airborne pollution concentration over the duration of contact with that air. However, the resulting dimensions of mass × time/volume may be difficult to conceive. Hence, some assessors express exposure as simply the concentration of pollutant in the contacted air, as a function of time. In this case, if the airborne exposure concentration is expresses in units of mass/volume, then that will also be the units of exposure, with the specification of that concentration as a function of time over a specified time interval or duration. Other assessors may use airborne concentration, multiplied by respiratory inhalation rate, as the expression of inhalation exposure. In this case, the dimensions of exposure would be mass/volume × volume/time, which equals mass per unit time. However, other assessors would call this product an applied inhalation dose rather than exposure. This discussion illustrates the sad fact that there are not consensus definitions of terms used in the field of exposure modeling.

1.2.3.2. List of human exposure models in THERdbASE

The following human exposure models are currently available in THERdbASE. The user can obtain information about each of these models by accessing the metadata for the model of interest.

1. Model 101: Subsetting Activity Pattern Data
2. Model 102: Location Patterns - Simulated
3. Model 107: Exposure - Inhalation - BEAM
4. Model 108: Exposure - Inhalation - Multiple Chemicals

1.2.4. Dose models

Dose refers to the amount of pollutant which enters the human body or specific organs or tissues of the human body. “Dose” has several different definitions, depending on the site of the pollutant within the body. Terms such as “potential dose,” “applied dose,” “internal dose,” “delivered dose,” and “biologically effective dose” are used to describe dose at various points along the route that pollutants follow as they enter the body and its organs. EPA’s Guidelines for Exposure Assessment (EPA 1992) contains definitions for these terms.

1.2.4.1. Overview

Dose models describe the amounts of pollutants which penetrate the human body and/or target organs within the body. Some dose models simulate the amount of pollutant delivered to the external boundary of the human body, such as the amount of an air pollutant passing into the nostrils during inhalation. Other dose models represent the body as a relatively simple system of a few interconnected compartments among which the pollutant distributes. Still other more complex dose models may be very elaborate representations of multiple organs and the blood circulatory system. These latter models are referred to as *physiologically-based pharmacokinetic* (PBPK) models. PBPK models are used to simulate the distribution of substances to various organs within the body, as well as the metabolism, storage, and elimination from organs and the body.

Because different dose models simulate differently defined doses, model users should check carefully to ascertain exactly what the operational model is computing. This information is usually included in the metadata for each model.

1.2.4.2. List of dose models in THERdbASE

The following dose models are currently available in THERdbASE. The user can obtain information about each of these models by accessing the *metadata* for the model of interest.

1. Model 109: Dose - Dermal - Film Thickness
2. Model 110: Scenario - Dose - Inhalation/Dermal
2. Model 201: Soil Exposure/Dose Assessment

1.2.5. Risk (health effects) models

Risk models are not a part of exposure modeling *per se.* Rather, risk models are used in conjunction with the results of exposure modeling or exposure assessment to estimate the likelihood that the estimated exposure will cause harm. Hence risk models are used in conjunction with exposure models to perform risk assessments.

1.2.5.1. Overview

There are generally two types of risk models used in assessing risk associated with environmental pollutant exposure. These are linear dose-response models, and threshold models.

In linear dose-response models, the risk or probability of a given adverse health effect is assumed to be directly and linearly proportional to the exposure (or possibly the dose). In this case, risk is conceived as a probabilistic phenomenon. A given exposure may or may not cause a specific adverse health effect. This type of risk model characterizes the statistical likelihood that a given exposure will cause a given effect. For example, cancer risk is generally assumed to be linearly proportional to exposure, as expressed by a *slope factor*. When this type of risk model is applied to a population exposure distribution, the result is a distribution of the probabilities of health effect in the modeled population.

Threshold models are generally used to estimate non-cancer health effects. This type of model assumes that low exposures below some threshold will have no adverse health effect, while exposures greater than the threshold will cause an effect.

1.2.5.2. List of risk models available in THERdbASE.

There are currently no risk models available in THERdbASE.

1.3. Applicability of HEMs

Human exposure modeling (HEMs) has several important applications. The first is to assess the exposure(s) to an existing pollutant for a given individual or an existing population or sub-population. This is an application for which the exposure model results can be evaluated by comparison to actual measurements of exposures among members of the modeled population.

Another application of human exposure modeling is the assessment of exposures for a potential new pollutant or change in emissions of an existing pollutant source. This scenario analysis is an important use of exposures models because, by definition, it estimates something that does not yet exist and therefore is not available for actual measurement. In cases like this, human exposure modeling is the only possible way of assessing exposures.

Another use of exposure modeling is in epidemiological studies. Models can be used to quantitatively estimate exposures of study subjects and controls who have or have not manifested certain health problems when an investigation is underway to try to determine causes of the health problems.

The latter two applications involve source apportionment, which is an important goal in many exposure assessments, and is sometimes an important objective for human exposure modeling.

Different types of exposure models may be used for different purposes or applications. For example, purely *stochastic* models may be used to portray existing exposures for which some measured data are available, but this type of model may not shed much light on the contributing sources of pollutant which cause those exposures. In order to evaluate the effects of sources, deterministic models which include explicit sources are likely to be more valuable. These different types of models are discussed more fully in the next section.

1.4. Mathematical aspects

Human exposure models may be classified by the type of method used to estimate exposures. The two major types are *deterministic* models and *stochastic* models. Models may also be classified according to the mathematical solution methods employed, which may be either *analytical* or *numerical approximation* methods. These mathematical aspects of models are discussed below.

1.4.1. Model types

Human exposure models may be classified, by type, on the basis of whether they perform purely deterministic calculations or whether one or more elements of the model is a random variable. The former are called *deterministic* models, and the latter are called *stochastic* models. Many models combine both and may be referred to as hybrid models.

1.4.1.1. Deterministic

Deterministic models have the characteristic that a model run with a given set of input data will always provide the same output. The model outputs are determined completely by the input; hence these models are called *deterministic*. Values of model input parameters are specified exactly, as opposed to being sampled randomly from a distribution of values. Thus, deterministic models generally provide a single value as a calculated output.

1.4.1.2. Stochastic

The word *stochastic* is derived from a Greek word meaning *to guess at.* In a stochastic model, one or more of the input parameters can accept multiple values which are randomly sampled from a distribution of possible values. Thus, each model simulation result is subject to random variation. Therefore, stochastic models are generally run for a relatively large number of iterations or *realizations*. The results of a number of realizations may then be organized into a distribution and considered collectively. Models run in this fashion are frequently called *Monte Carlo* models or *probabilistic* models.

Several of the models in THERdbASE are set up as stochastic models. The *Benzene Exposure Assessment Model (Model 107: Exposure - Inhalation - BEAM)* is one such model. This model computes exposure by stochastically sampling concentrations of benzene from one of several distributions and multiplying exposure duration and inhalation rate by the sampled concentration.

Stochastic models are most useful when used to simulate a characteristic which is difficult or impossible to determine with a high degree of certainty for specific instances, but about which we know the statistical distribution properties. *Human activity patterns* are a good example. We may not be able to accurately predict or simulate (model) where a given individual person will be at a given moment of the day, but we may know from population studies what the probability is of people being in a particular type of location. Likewise, we may not be able to accurately predict a pollutant concentration at a given place at a given time, but we can measure or model the statistical distribution of pollutant concentrations at various types of locations over time. In a stochastic model, then, we can sample from a distribution of activity patterns to determine times spent in various locations and multiply those durations by concentrations sampled from a distribution of concentrations relevant to that type of location, and thus model the population’s distribution of exposures. The statistical properties of the resultant distribution may be quite accurate, even though we may know nothing about the activities or exposure of any specific individual in that population.

Stochastic models are widely used in human exposure modeling because several of the key variables upon which exposure depends are *random variables* rather than deterministic variables which can be accurately predicted for individual cases. Human activity patterns provide a good example.

One of the primary uses of running models stochastically, i.e., using Monte Carlo samples, is to characterize the variability or uncertainty in a model output. In situations where this is desired, Monte Carlo modeling may be very useful because the output is itself a distribution showing the expected probabilities of a range of results. This provides the modeler with considerably more information than would a single-value output from a deterministic model. EPA has published guidance on the use of Monte Carlo analysis in exposure assessment (EPA 1997).

1.4.1.3. Hybrid

Hybrid deterministic-stochastic models are those which contain both deterministic and stochastic elements. Many exposure-related models contain both types of parameters and hence may be considered hybrids. A good example would be a model such as the BEAM mentioned in the preceding subsection, in which one or more of the variables was kept at a fixed value while another variable is varied randomly by Monte Carlo sampling. Such a modeling run could be part of a sensitivity analysis, in which the model's output sensitivity to its input values is analyzed. When the model is run in such a mode, the model is a hybrid since some parameters are fixed and others are varied stochastically.

1.4.2. Solution methods

Computer programs can use two general methods to solve equations within a model - analytical or numerical solution techniques. The following subsections briefly describe these methods.

1.4.2.1 Analytical

When equations are said to be solved analytically, an exact mathematical solution is found for the equations. For example, a differential equation may have an exact or analytical solution computed by the software.

1.4.2.2. Numerical

Many exposure-related models contain differential equations or systems of differential equations for which no analytical solution is known, or for which the execution of an analytical solution may be very computationally difficult. In such cases, numerical approximation routines commonly obtain approximate solutions to the equations. A typical example is a multi-compartmental indoor air model (e.g., Model 106: Indoor air - N Zones). Each of several rooms within a building is represented by a mass-balance differential equation. The resulting system of simultaneous differential equations is solved by a numerical integration procedure which steps the equations through a series of very small increments of the dependent variable (time) and estimates from the discretized differential equations the change in all system variables after each small time step. An explicit analytical expression or equation characterizing the values of the dependent variables is not found; rather the program computes an estimated numerical solution to the equations at each time step.

1.5. Inputs

All models need data inputs before they can be run. Most models incorporate several or many factors whose values can be varied by the user. The data values provided by the user are referred to as inputs to the model. As described in this section, model inputs can be provided in several forms, described below.

1.5.1. Default constants

A default constant is the simplest way to specify an input value for a model. A default constant is a single, pre-programmed numerical value which the model uses as the input value for a parameter.

For example, assume that a model calculates inhalation dose by multiplying airborne pollutant concentration to which a person is exposed by the person's respiratory inhalation rate and by the duration of exposure. The model can be expressed as follows:

Exposure = concentration × inhalation rate × duration.

For this simple model, the input parameters are concentration, inhalation rate, and duration. If the model provides a pre-programmed default value of 10 liters per minute for the inhalation rate, then 10 liters per minute is the default constant for that input.

The user should be cautious when using models which provide default constants as model inputs. The value of the default constant may not be appropriate or applicable for the situation being modeled by the user. In the example above, 10 liters per minute may not be a good estimate of the inhalation rate of the person being modeled. Users should always check the default constants used in a model, and if they are not appropriate, change them to appropriate values before running the model.

The metadata provided with each model should contain a brief statement about the default constants and why they were chosen by the model developer. Please also review the valuable information about THERdbASE models in *Appendix A - Model Notes.*

1.5.2. User-input constants

As noted in the preceding paragraph, models sometimes may contain default constants as inputs. When a provided default constant is not appropriate for a given use, the user may provide a different single numerical value in place of the default constant. Such an input is referred to as a user-input constant.

Some models are programmed to allow only user-input constants that fall within a specified range. An attempt by the user to input a value which is outside the allowable range may cause a cautionary warning note to appear on the computer monitor. The metadata provided with each model should contain information about allowable ranges for input parameters. (Also see Appendix A - Model Notes.)

1.5.3. Mathematical distributions

The two preceding sub-sections defined constant-value inputs to models. In the cases described above, the same value (either a default constant or a user-input constant) is used each time the model executes, until the user changes the value. Hence those input values are termed *constant* values. However, there are many times when a modeler wants to run a model stochastically to determine a distribution of results. In these cases, an input parameter can be specified to have a different value each time the model executes. The value for the input parameter can be randomly selected from a distribution of values during each model execution.

There are several ways, described briefly in the following subsections, that the sampling of a value from a distribution can be accomplished. The sampled distribution can be either a mathematically defined distribution, or an actual data distribution.

In sampling from a mathematically defined distribution, the values to be used as input parameter values are created by a mathematical routine within the software. *There are many mathematically defined distributions, such as the uniform distribution, the normal distribution, the log-normal distribution, the beta distribution, and the Weibull distribution*. Each distribution can be mathematically defined by two or more parameters, such as the mean and standard deviation of a normal distribution. When the program needs an input value a random number is generated by the software. That random number is then used together with the distribution parameters to calculate a number from the distribution.

When this procedure is repeated for a sufficiently large number of iterations, the resulting distribution of values has the statistical properties as defined by its parameters.

The following sub-sections briefly describe the three types of mathematical distributions that are used in the exposure-related models within THERdbASE.

1.5.3.1. Uniform distribution

As its name implies, the uniform distribution is a set of numbers within a specified range, in which each number has an equal or uniform probability of occurring. A uniform distribution is specified by two parameters - a minimum and a maximum. Each possible number between the minimum and maximum value has an equal probability of being drawn from the distribution.

1.5.3.2. Normal distribution

The normal distribution has been found to be a close statistical approximation to many observed natural phenomena. For example, the measured heights of a population of humans will generally be found to have a normal distribution.

The normal distribution can be defined by a mathematical expression which has the following form: $f = e^{-z^2/2} / \sqrt{2\pi}$, where f is the probability density function of the *standard normal random variable*. This is the normal distribution which has a mean of zero and a standard deviation of 1.0 (Devore 1986). The parameter z in the above equation represents the variable which is normally distributed. This variable can have any value between negative infinity and positive infinity. The calculation *f* is the probability with which a given value of *z* occurs in the distribution.

When a parameter is normally distributed, a graph of its probability versus the parameter value will show the familiar bell curve. This is illustrated in *Figure 2* below.

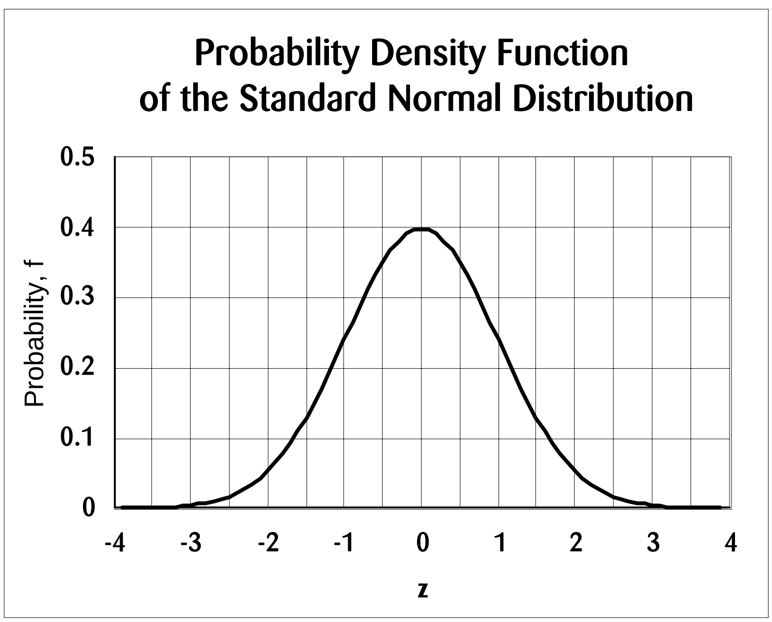

Figure 2. Graph of the Standard normal distribution.

When the value of a normally distributed variable is plotted on *probability paper*, the resulting curve is a straight line. Probability paper is a form of non-linear graph paper that has cumulative probability (also called cumulative frequency) on the non-linear axis and the parameter value on the linear axis. The graph shows the frequency with which values equal to or less than a given value will occur in the distribution. This is illustrated in *Figure 3* below.

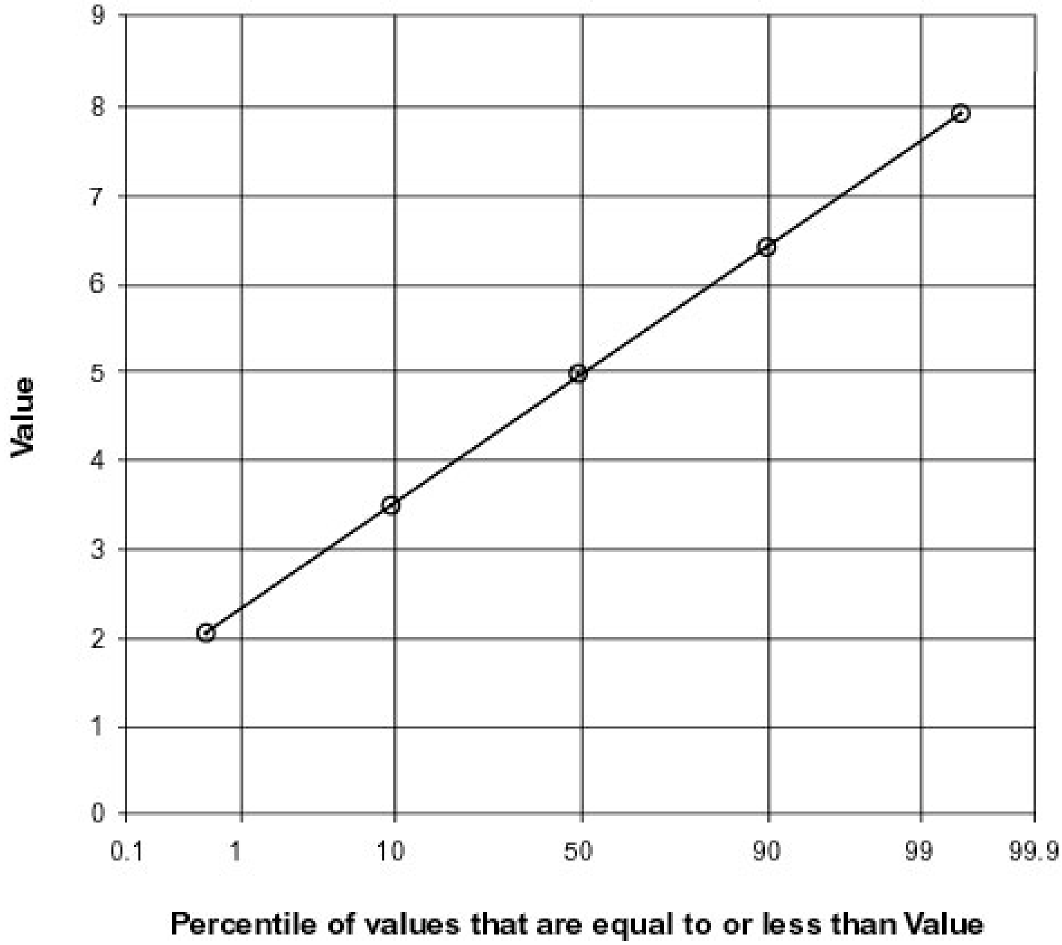

Figure 3. Cumulative frequency plot of a normally distributed variable.

One useful way to evaluate data to see if they are normally distributed is to create a plot of the data on probability paper. If the resulting plot is a straight line we may conclude that the data are normally distributed.

1.5.3.3. Log-normal distribution

When the natural logarithms of a set of data are normally distributed, we say that the data are "lognormally distributed". Many environmental variables are observed to be lognormally distributed. A physical explanation for why this is the case has been published (Ott 1990). In Figure 4 below, some exposure data (personal-air concentrations of a pollutant) from THERdbASE are plotted on log-probability paper. On this paper, the horizontal axis shows the cumulative frequency of values falling at or below a given value. Note that the vertical axis is logarithmic rather than linear. The fact that the example data fall along a nearly-straight line means that we can say that these data are "log-normally distributed".

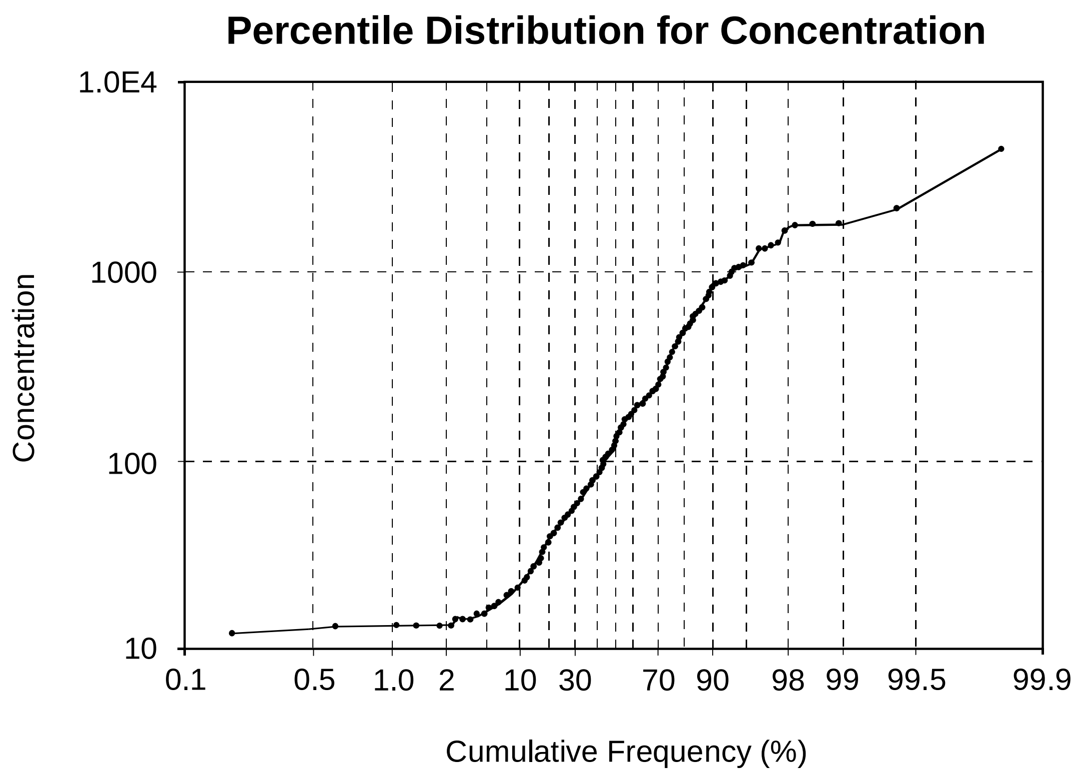

Figure 4. Cumulative frequency plot of concentration data, on log-normal probability coordinates.

1.5.4. Data distributions

Sometimes in exposure modeling we want to sample from actual data points in a list or distribution, rather than sampling from an abstract, mathematically defined distribution such as the normal distribution described above. In modeling software that samples from data distributions, the software generates a random number that represents a percentile on the data distribution. The software then retrieves, from an ordered list of values, the data point that corresponds to the randomly drawn percentile value. 1Some of the models in THERdbASE use this method, i.e., when a data value is needed in the model, the software executes a routine that selects an appropriate value from a distribution of data stored in a datafile.

1.6. Outputs

Models can produce outputs in a variety of formats, including tabular and graphical. For example, if the output of a model is the concentration of an air pollutant as it varies over time, the results might be available as a tabular list of times and their corresponding concentrations (e.g., concentration every hour over a 24-hour day). Frequently, modeling software also has the ability to provide the output in the form of a graphical plot. In the example just noted, the output might be in the form of a graph with time of day on the horizontal x-axis, and the corresponding air concentrations plotted on the vertical y-axis. Models which perform Monte Carlo simulations usually can provide outputs in the form of probability plots of the modeled distributions, such as the example shown in Figure 4.

Frequently, the outputs from modeling software can be either printed on paper, or saved to a computer file. Some modeling software will allow printing or saving information about the model input parameters and run conditions, as well as the computed results of the modeling. This information about the modeling run can be an important form of documentation for maintaining good records of the performance of the model. In THERdbASE, all the models are capable of printing or saving a variety of outputs, including information about the modeling run.

1.7. Interactions among models

In some complex modeling work, several different models are used in a way such that the output from one model becomes the input to another model. For example, a source model might simulate the rate of release of a chemical into the air. This source model output could then be used as an input into an indoor air concentration model. The later model might compute the concentration-versus-time profile of the pollutant in the air in a room. This concentration could in turn be used as the input to a human exposure model. When the person is in the indoor room simulated by the indoor air model, the corresponding air pollutant concentration would become an input to the model which computes the person's exposure to that pollutant.

The interactions among different models requires that the models be coded in an appropriately compatible way, so that the proper data are passed from one model to the other in a proper format and at the proper time. Attaining the needed degree of coordination among the various models is a programming challenge. The work involved in linking models is one of the major obstacles to be overcome in a complex modeling exercise such as human exposure modeling. In THERdbASE, the interactions among models, where it occurs, is done automatically by the software, i.e., the user does not have to manipulate one model's output to make it available to another model.

1.8. Model processing

After a user has set up a model for running by specifying all the needed input data and run conditions, the model is ready to be run. When the proper command is given, the computer begins processing the data and computing results. Depending on the complexity of the model, the number of simulations, and the computer operating speed, the model run will take a period of time to be completed.

1.9. Evaluations

Models can be evaluated by comparing their computed output results to some other data that are known by the user to be reliable. For example, if a model simulated an air pollutant concentration-versus-time profile, the model output could be evaluated by comparison to accurately measured air pollutant concentrations from a laboratory experiment whose conditions were the same as those represented in the model simulation. However, the user of a model or model output should be cautious to not read too much into the model results. Even though a model may give an accurate simulation of one or more specific data sets, this does not ensure that the model would be able to accurately simulate data collected under other conditions. Thus a model should be considered "valid" only for conditions for which it is known by the user to be accurate. No model can be "validated" universally. In other words, we should not trust models to be able to accurately simulate reality other than for the conditions on which the model's development was based. Model results should always be viewed with a discerning and skeptical eye. Remember the words of the statistician George Box: "All models are wrong but some models are useful."

1.10. Glossary

For further definition and explanation of many of the terms used in this tutorial, please refer to the Glossary available within THERdbASE.

1.11. Contacts

Any of the following persons may be contacted for further information about using THERdbASE for exposure modeling:

Phone Numbers
(702) 798-2664 for William H. Engelmann
(702) 798-2285 for Edwin J. Furtaw
(702) 798-2361 for Joseph V. Behar
(702) 798-2114 for Larry C. Butler

Fax number
(702) 798-2261

1.12 E-mail and Affiliation

engelmann.william@epa.gov
furtaw.ed@epa.gov
behar.joseph@epa.gov
butler.larry@epa.gov

U.S. Environmental Protection Agency
Office of Research and Development
National Exposure Research Laboratory
Human Exposure and Atmospheric Sciences Division
Human Exposure Research Branch
P.O. Box 93478
Las Vegas, NV 89193-3478